An Adaptive Retransmission Scheme with QoS Support for the IEEE 802.11 MAC Enhancement

Wen-Tsuen Chen Bo-Bin Jian Shou-Chih Lo
Department of Computer Science
National Tsing Hua University
Hsinchu, Taiwan 300, R.O.C.
Tel: +886-3-5742896 Fax: +886-3-5711484
Email: jbb@cs.nthu.edu.tw

Abstract- **The Medium Access Control (MAC) protocol of the IEEE 802.11 standard is based on Carrier Sense Multiple Access with Collision Avoidance (CSMA/CA). The basic retransmission mechanism, binary exponential backoff, may cause large packet delay and jitter values that are not suitable for real-time traffic. In this paper, we first investigate some MAC enhancement mechanisms discussed in the IEEE 802.11 task group E, which was formed for enhancing the current 802.11 MAC protocol to support for applications with Quality of Service (QoS) requirements. Then, we propose a jamming-based retransmission mechanism that is compatible with the 802.11 standard and could reduce the packet delay of real-time traffic. Besides, this mechanism performs stably when the traffic load is heavy. The optimal setting of our proposed mechanism is discussed analytically. We perform simulated experiments by comparing our proposed retransmission mechanism with the other two mechanisms discussed in the 802.11 task group E. We show that the jamming-based retransmission mechanism can reduce the packet delay and the packet dropping rate.**

I. INTRODUCTION AND BACKGROUND

As the mobile devices and Internet services become popular, the trend on providing wireless data services has emerged. The great improvement in channel modulation techniques makes the bandwidth of wireless medium large. It becomes feasible for real-time multimedia data to be transmitted via the wireless medium.

The IEEE 802.11 Wireless Local Area Network (WLAN) [14] is an international standard which specified the Medium Access Control (MAC) sub-layer and Physical (PHY) layer. The IEEE 802.11b using the 2.45GHz ISM band can provide 11Mbps data rate and the 802.11a operating on the 5GHz radio frequency with the OFDM modulation scheme can offer up to 54 Mbps data rate [2]. Therefore, the WLAN with high speed and low cost access to the Internet is a good platform to provide real-time services.

The basic building block of an IEEE 802.11 WLAN is called basic service set (BSS), which is a set of wireless stations (STA) controlled by a coordination function. The 802.11 WLAN can be configured as an ad hoc network (an independent BSS) or an infrastructure network (composed of an access point and the associated STAs).

In the 802.11 WLAN, the channel access for the STAs in a BSS is under the control of two types of coordination functions: Distributed Coordination Function (DCF) and Point Coordination Function (PCF). The fundamental access method is DCF, which is based on the CSMA/CA technology. The CSMA/CA is a contention-based multiple access technology that requires each STA to sense the medium to be idle for a period of time before sending each frame. The period of time is called inter-frame space (IFS) whose length is related to the frame priority. The levels of frame priorities are classified as Short IFS (SIFS), PCF IFS (PIFS), DCF IFS (DIFS) and Extended IFS (EIFS), which correspond to, for example, ACK frames, PCF control frames, data frames, and retransmission frames, respectively. More details about the frame priority could be referred to [1]. In the DCF, the backoff procedure is used for collision avoidance, where each STA waits for a backoff time (a random time interval) before its frame transmission.

The backoff procedure starts once a STA detects the medium busy and attempts to transmit a data frame. The STA will calculate a random backoff time in units of slot-times according to the Contention Window (CW). The CW specifies the range of possible backoff time. When a collision occurs during the frame transmission, the CW becomes double and is bounded by a maximal value, CWmax. This is so called binary exponential backoff in the DCF. The STA will decrease its backoff time counter by one if the medium is sensed idle for a slot-time and freezes the backoff time counter when the medium is busy. When the backoff time is decreased to zero, the STA transmit its frame immediately. The generating of random backoff time can be described as (1). The RandomInteger(Minimum, Maximum) function generates a random integer in the range from Minimum to Maximum uniformly.

$$BackoffTime = RandomInteger(0, \min(CWmin \times 2^{retries}, CWmax)) \times SlotTime \quad (1)$$

With the binary exponential backoff procedure, the backoff time may increase tremendously for each retransmission. It causes a large packet delay. Moreover, the changes of CW values make the packet jitter large, especially after a successful retransmission and the increasing CW is immediately reset to the minimal value, CWmin. Hence, the binary exponential backoff is not suitable for the transmission of real-time traffic. To enhance the current IEEE 802.11 MAC to support for applications with Quality of Service (QoS) requirements, the IEEE 802.11 task group E is formed and is proceeding to build the QoS enhancement of the MAC protocol. The initial 802.11E draft is currently coming up [7]. The enhancement mechanisms for the DCF access discussed in the 802.11 task group E leave the same retransmission problem for real-time traffic unsolved. In this paper, we propose an efficient and adaptive retransmission mechanism using jamming noises to resolve the contention between real-time traffic and separate the

0-7803-7484-3/02/$17.00 ©2002 IEEE.

contention of real-time traffic from non-real-time traffic. Also, we study the generation of jamming noises analytically.

The remainder of this paper is organized as follow. Section II gives a survey on related researches about 802.11 MAC enhancements for real-time traffic. Section III describes our proposed jamming-based retransmission mechanism and discusses the optimal parameter setting of our proposed mechanism. In Section IV, we show the performance evaluations through simulation results. Our conclusions and future work are presented in Section V.

II. RELATED WORK

Providing QoS guarantees becomes an important issue nowadays [3]. Some polling based mechanisms were proposed to serve real-time traffic [4][5]. In [4], a variation of EDF (early deadline first) scheduling for real-time traffic and a variation of Round-Robin scheduling for non-real-time traffic are used. In [5], real-time traffic and non-real-time traffic are served by the PCF and the DCF, respectively. The differentiation services are provided in the DCF by using a "rollback" backoff mechanism. Four different traffic classes: conversation, streaming, interactive and background proposed by UMTS [6] are used. However, the implementation complexity of polling functions may be high.

Many proposals addressing the DCF enhancements were discussed in the IEEE P802.11 task group E. [8] summarized the features of these enhancements based on six functions: basic contention resolution approach, class differentiation, packet differentiation, averting packet aging, scheduling of competing traffic streams, and adaptation to traffic intensity. It also indicates the issue about the coexistence between these features. In [9], a history review of the proposals for DCF enhancements is given.

The Virtual DCF (VDCF) mechanism [10][11] with the low implementation cost was discussed most popularly. The idea of VDCF is to adjust the sizes of CW and IFS according to the traffic priority. The real-time traffic with a smaller setting obtains more opportunities to access the channel. Each flow has its own backoff time setting inside a STA. Conceptually, the VDCF has a parallel queue for each traffic catalog and contentions between these queues will happen first inside the STA. However, the retransmission mechanism, which is also the binary exponential backoff, still cause large packet delays for real-time traffic and the situation would be worse when the channel condition is bad or the traffic load becomes heavy. It is unfavorable for real-time traffic with time-bounded requirements.

Another scheme, named Tiered Contention Multiple Access (TCMA) [12][13], uses a smaller CW for each retransmission. The TCMA scheme maps eight traffic categories into four Urgency Classes (UC) according to the IEEE 802.1d Annex H.2. Each UC has it's own backoff time setting. Packets queued in certain classes will aging, and if the age exceeds the threshold of the class, the UC of these packets will be dynamically updated to reflect their "transmission urgency". The retransmission mechanism in the TCMA was different with others. It decreases the CW for each retransmission according to a scaling parameter, CWPFactor. The CWPFactor is set according to the traffic priority. The relationship between the current CW and the new CW is shown in (2). The diminished CW is used for reducing the overhead of backoff time. Nevertheless, this scheme incurs more packet drops due to enormous collisions in a heavy traffic load as will be shown in our experiments.

$$CW_{new} = \lceil (CW_{current} + 1) \times (CWPFactor / 16) \rceil - 1 \qquad (2)$$

III. THE ADAPTIVE RETRANSMISSION SCHEME

We first show the frequency of occurring retransmissions in the CSMA/CA with binary exponential backoff scheme through a simple simulation with various numbers of STAs. The result is shown in Fig. 1. As can be seen, large retransmission reties happen frequently when there has a large amount of contending STAs. The long-round frame retransmission will be a serious problem for time-bounded and jitter-sensitive real-time traffic. If the channel condition is bad too, the retransmission problem becomes more serous.

A. Jamming-Based Retransmission Mechanism

In this paper, we propose an efficient and adaptive retransmission mechanism, named jamming-based retransmission mechanism, which can limit the packet delay and the jitter for real-time traffic. From the above experiment, we know that serious collisions imply either a heavy traffic load or a large amount of retransmission packets. If we can guarantee the retransmission packets to be transmitted with a proper control, they will not be involved in the serious contentions so as to improve the overall system performance. In our scheme, we separate non-real-time traffic from contenting with real-time traffic and let real-time flows transmit their retransmission packets as soon as possible.

Suppose that the retransmission of a real-time frame starts a jamming procedure instead of a backoff procedure to access the channel. When the channel is idle for a DIFS period, STAs with retransmission real-time frames will generate jamming noises and continuously send the signal for a period of random time. A STA starts frame retransmission if the channel remains idle after sending its own jamming noise. Basically, the jamming procedure will block those STAs performing backoff procedures from contending for the channel. At least one STA will survive the jamming phase and the jamming retransmission will be successful if there is only one survival. Throughout, a STA with the support of the jamming function is called enhanced STA (ESTA). Fig. 2 depicts the situation when four ESTAs, ESTA1 to ESTA4, are performing the jamming procedures and a legacy STA or an ESTA with non-real-time retransmission frames is performing the backoff procedure. ESTA 1 with the longest jamming noise survives and others (ESTA2-ESTA4) wait for the next transmission opportunity. The last STA will freeze its backoff time counter because the channel is always sensed busy. The collision occurs only when more than one ESTA has the same longest jamming time period. It is most likely to have a successful retransmission in each transmission opportunity with a proper control of jamming time.

The length of jamming time is a random number in units of slot-times. The probability of a jamming time with f slot-times long, $P_J(f)$,is given by the truncated geometric distribution in (3). p_J is a

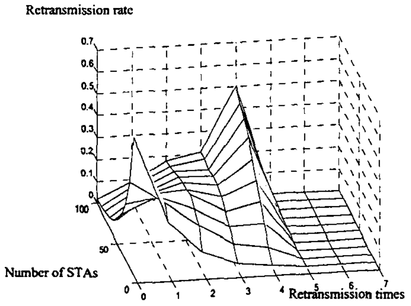

Fig. 1. Retransmission rate.

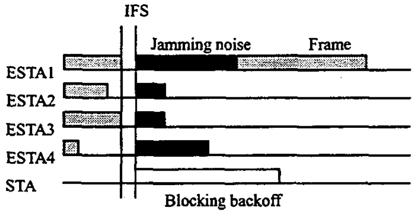

Fig. 2. Jamming-based channel contention.

TABLE I
RECOMMENDED *JW* VALUES

p_J	*JW*	p_J	*JW*	p_J	*JW*
0.05	4	0.40	10	0.75	28
0.10	4	0.45	11	0.80	35
0.15	5	0.50	13	0.85	45
0.20	6	0.55	15	0.90	66
0.25	7	0.60	17	0.95	122
0.30	8	0.65	19		
0.35	9	0.70	23		

probability parameter between 0 and 1, and *JW* (Jamming Window) is defined as the maximum jamming time. The probability distribution is more skewed with a smaller p_J and it is most possible for an ESTA to survey with a short jamming period. The setting of *JW* is mainly dependent on the number of ESTAs simultaneously performing the jamming procedures, which is denoted as *N*. The simple criterion in (4) can be used to decide the proper *JW* value, where a large *JW* value is used to reduce the collisions on the same jamming time for a large *N* value. For example, the *JW* value is 6, 10, 17 and, 35 when p_J value is 0.2, 0.4, 0.6, and 0.8, respectively as N=1000. Table I collects the recommended *JW* value with various p_J values.

$$P_J(f)=\begin{cases} p_J^{f-1}\cdot(1-p_J), & 1\le f<JW \\ p_J^{JW-1} , & f=JW \end{cases} \qquad (3)$$

$$JW=\min\left\{JW \mid p_J^{JW-1}\le\frac{1}{N}\right\} \qquad (4)$$

B. Parameters Optimization

We analyze the jamming retransmission cost to decide the proper p_J value. Assume a transmitted data frame has the average size of *framesize* octets. Let *RetryLimit* denote the maximum number of retransmission attempts. The probability of *i* ESTAs generating the longest jamming time of *f* slots is given in (5). The average retransmission costs for a successful jamming and an unsuccessful jamming are expressed in (6). Therefore, we obtain the average retransmission cost, *Ravg*, as shown in (7). We can determine the optimal p_J value with the minimum retransmission cost, given the size of a data frame and the number of ESTAs. For example, the optimal p_J value is 0.4, 0.5, and 0.55 when data *framesize* is 200, 300, and 400 octets, respectively as *N* is 1000. Fig. 3 (a) shows the numerical result of average retransmission cost as *RetryLimit* is 5. The upper curved surface is with N = 1000 and the lower one is with N = 100. It can be seen that the retransmission cost is insensitive to N. Fig. 3 (b) shows the optimal p_J value with the minimum retransmission cost, given the size of a data frame. Therefore, we first decide the p_J value from the size of the data frame, and then decide the *JW* value.

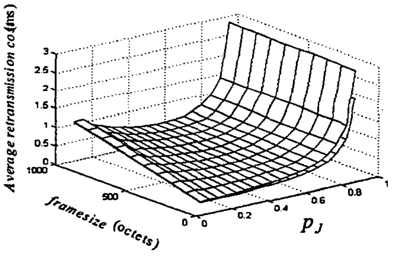

Fig. 3. (a) Average retransmission cost.

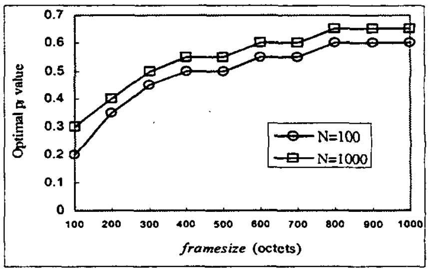

Fig. 3. (b) Optimal p_J value.

$$P_J(N,f,i)=\binom{N}{i}P_J(f)^i\left(\sum_{j=1}^{f-1}P_J(j)\right)^{N-i} \quad (5)$$

$$\begin{aligned}
&R_{success}\ (\textit{a successful retransmission cost}) = \\
&\quad JAM_{success}\ (\textit{average jamming time} \mid \textit{jamming is successful}) \\
&R_{failure}\ (\textit{an unsuccessful retransmission cost}) = \qquad (6)\\
&\quad JAM_{failure}\ (\textit{average jamming time} \mid \textit{jamming is unsuccessful}) + \\
&\quad T_{framesize}\ (\textit{time to tramsmit a data frame}) \\
&JAM_{sucess} = \sum_{f=1}^{JW} P_J(N,f,1)\cdot f \\
&JAM_{failure} = \sum_{f=1}^{JW}\sum_{i=2}^{N} P_J(N,f,i)\cdot f
\end{aligned}$$

$$R_{avg} = \sum_{i=0}^{RetryLimit} \{ \left(i\cdot R_{failure} + R_{success}\right)\cdot \left(1-\sum_{f=1}^{JW} P_J(N,f,1)\right)^i \cdot \sum_{f=1}^{JW} P_J(N,f,1) \} \quad (7)$$

IV. SIMULATION RESULTS

We do the performance evaluation through the simulation program. First, we evaluate the performances by assuming each STA always has a pending data frame to be transmitted. Then, we use certain traffic models to perform the evaluations again. The channel capacity is 10 Mbps and the channel condition is assumed to be error-free. The real-time data offered load is defined as the number of active real-time traffic flows. Seventy non-real-time traffic flows act as the background traffic. Each ESTA can deal with one traffic flow at the same time, and the flows always attempt to access the channel in each transmission opportunity. The *framesize* is 160 octets for real-time traffic and is 512 octets for non-real-time traffic. Here, we compare the three schemes: VDCF, TCMA, and our proposed jamming-based scheme (Jamming), which use the same access method except for the retransmission mechanism. The (CWmin, DIFS) values are (15, 40us) and (31, 50us) for real-time traffic and non-real-time traffic, respectively. The CWPFactor is 8 in the TCMA. p_J is 0.35 and JW is 9 in the Jamming. We compares between these three schemes from the aspects of mean MAC delay (in millisecond), packet dropping rate, average jitter and channel utilization when real-time traffic flows increase. The mean MAC delay is the time period from the start contending for the channel to the end of successful transmission. The average jitter is defined as the average difference between previous and current successful frames delay. The channel utilization stands for the rate of successful data transmission time.

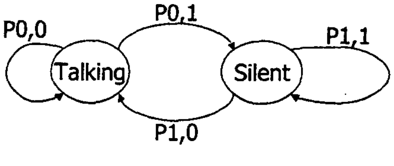

Fig. 4. Voice traffic model by a two-state Markov chain.

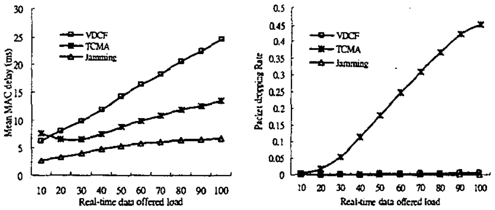

Fig. 5. (a) (b)

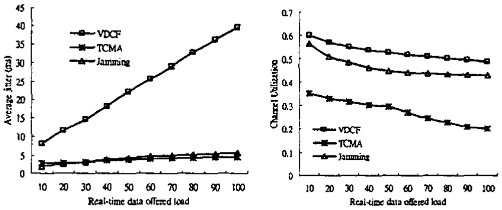

Fig. 5. (c) (d)

Fig. 5. Comparison between different retransmission mechanisms.
(a) Mean MAC delay. (b) Packet dropping rate.
(c) Average jitter. (d) Channel utilization.

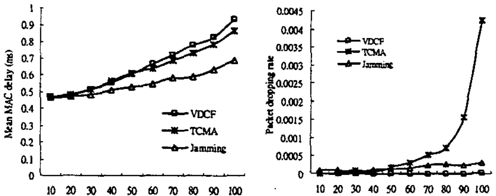

Fig. 6. (a) (b)

Fig. 6. Comparison between different retransmission mechanisms with the traffic model of two-state Markov chain.
(a) Mean MAC delay. (b) Packet dropping rate.

In Fig 5. (a)(b), the Jamming has a lower MAC delay than other two schemes and the MAC delay is less sensitive to the traffic loads. Notice that the TCMA has a little more delay when the real-time traffic offered load is low because the packets do not exceed the retry limit. With the increasing offered load, the packet dropping rate becomes huge and delay time increases because collisions are aggregated with a smaller CW. TCMA quickly drops the packets for lower delay. It is not a good way to reduce the delay because the enormous dropping frame may cause the upper-layer of the network to retransmit these packets for reliability and it may aggravate congestion in the MAC layer. Even though no retransmission in the upper layer, it is still hard to match the boundary of dropping rate for real-time traffic. This experiment reveals the ability of the jamming retransmission to limit the packet delay. The stable performance is attributed to the feature for our jamming rule: a large probability to survive a station in each transmission opportunity no matter how many stations is involved in the contention.

Fig. 5. (d) shows the Jamming improves 36.48% channel utilization than TCMA, but 12.25% less than VDCF. The reason is that VDCF gives more chance to non-real-time traffic and Jamming wastes a little jamming time. The outcome reflected in the jitter of real-time traffic is shown in Fig. 5. (c). VDCF is from 3.08 to 6.09 times the average jitter of Jamming.

Furthermore, we design a simulator by the C language. The physical parameter is followed IEEE 802.11b specific DSSS mode. Refer to [15] which discussed many traffic characterization and simulation analysis, we use voice traffic as the real-time traffic that is modeled by a two states Markov chain as shown in Fig 4. The traffic of two different flows is assumed to be independent. A transition between the 'Talking' and 'Silent' states can happen at anytime with the transition probability Pt,s (Ps, t) from the 'Talking' to 'Silent' (from 'Silent' to 'Talking'). For a given average time duration Tt (Ts) at 'Talking' ('Silent') state Pts (Pst) is exponential distributed with mean 1/Tt (1/Ts). Here we assume Tt=1000 and Ts = 1350 (millisecond). The data rate is 64Kbps and frame duration is a fixed 20 millisecond. The non-real-time traffic is generated by an ON/OFF source with ON state duration to be 3.3 sec and OFF state duration is 22.8 sec. The ON and OFF periods are distributed according to Weibull distribution. The *framesize* is bimodal distributed from 256 to 512 bytes. The simulation result is shown in Fig 6. The curve is matched with our previous observation.

V. CONCLUSION AND FUTURE WORKS

In this paper, we proposed a jamming-based retransmission mechanism, which can guarantee the delay bound of real-time traffic in the contention period. With the jamming function, the real-time traffic could be rapidly transmitted without the effects of large delay and variant jitter which may be caused by the binary exponential backoff procedure specified in the IEEE 802.11 MAC. Our proposed mechanism can achieve a stable performance regardless of the traffic loads. We have used an analytical method to determine the proper length of a jamming noise, which is mainly dependent on the size of data frame being transmitted. The STAs with the jamming function are compatible with legacy STAs. The proposed retransmission mechanism can be combined with other internal packet scheduling schemes to achieve a better performance. In the future, we will study the differentiation of jamming retransmissions according to the traffic priorities. The call admission control will be involved too.

ACKNOWLEDGMENT

This research was partially supported by the Educational Ministry of the Republic of China under Contract No. 89-E-FA04-1-4.

REFERENCES

[1] ISO/IEC 8802-11; ANSI/IEEE Std 802.11, 1999 edn , 20 Aug. 1999 "Wireless LAN Medium Access Control (MAC) and Physical Layer (PHY) specifications".

[2] ISO/IEC 8802-11:1999/Amd 1:2000(E); IEEE Std 802.11a-1999 "Supplement to IEEE standard for Information technology: High-speed Physical Layer in 5 GHZ Band".

[3] I. Aad and C. Castelluccia, "Differentiation Mechanisms for IEEE 802.11," *Proc. IEEE INFOCOM*, vol. 1, pp. 209-218, 2001.

[4] Sunghyun Choi and Kang G. Shin "A Unified Wireless LAN Architecture for Real-Time and Non-Real-Time Communication Service," *IEEE/ACM Transaction on Networking*, vol. 8, No 1, Febuary 2000.

[5] Wen-Tsuen Chen, Jun-Jia Dai, and Shou-Chih Lo, "A MAC Protocol with QoS Guarantees for Multiclass Traffics in Wireless LANs," *Proc. The 13th International Conference on Wireless Communications (Wireless 2001)*, July 9-11 2001, vol. 2, pp. 450-459.

[6] 3GPP, "3rd Generation Partnership Project; Technical Specification Group Services and System Aspects; QoS Concept and Architecture (Release 1999)".

[7] IEEE P802.11 - TASK GROUP E, "http://www.ieee802.org/11/".

[8] M. Benveniste, "Summary of Features Proposed for the Enhanced-DCF Wireless MAC Protocol," *IEEE Document, 802.11-01/003*, January 2001.

[9] Jin-Meng Ho, Khald Turki and Sid Schrum, et al, "ECA Overview," *IEEE Document, 802.11-01/086*, January 2001.

[10] G. Chesson, A. Singla, W. Diepstraten, and H. Tenniusen, et al, "VDCF State Machine Description," *IEEE Document, 802.11-00/412r1*, November 2000.

[11] G. Chesson, A. Singla, W. Diepstraten, and H. Tenniusen, et al, "VDCF Proposed Draft Text," *IEEE Document, 802.11-01/131*, March 2001.

[12] M. Benveniste, "Proposed Normative Text for TCMA with Backoff Adaptation," *IEEE Document, 802.11-01/135r1*, March 2001.

[13] M. Benveniste, "Overview of 'Tier Contention' Multiple Access (TCMA)," *IEEE Document, 802.11-01/019*, January 2001.

[14] B. P. Crow, I. Widjaja, J. G. Kim, and P. T. Sakai, "IEEE 802.11 Wireless Local Area Networks," *IEEE Communications Magazine*, vol. 35, Issue 9, pp. 116-126, September 1997.

[15] G. Anastasi and L. Lenzini, "QoS provided by the IEEE 802.11 wireless LAN to advanced data applications: a simulation analysis," *Wireless Networks, ACM Journal*, vol. 6 issue 2, pp.99-108, 200.